The Patent Office

PCT/GB 98/02259

09/463844

28 July 1998

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP9 1RH

REC 31 AUG 1998
WIPO PCT

PRIORITY DOCUMENT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Andrew Gerry

Dated 10th August 1998

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
B
-1 AUG 1997
RECEIVED BY POST

The Patent Office

01AUG97 E292956-1 D02934
P01/7700 25.00 - 9716162.4

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

   PHM 97-043

2. Patent application number
   (The Patent Office will fill in this part)

   9716162.4

   01 AUG 1997

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

   ZENECA Limited
   15 Stanhope Gate
   LONDON, W1Y 6LN, Great Britain

   Patents ADP number *(if you know it)*

   6254007002

   If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention

   NOVEL COMPOUNDS

5. Name of your agent *(if you have one)*

   PHILLIPS, Neil Godfrey Alasdair

   "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

   Intellectual Property Department
   ZENECA Pharmaceuticals, Mereside
   Alderley Park, Macclesfield, Cheshire, SK10 4TG, Great Britain

   Patents ADP number *(if you know it)* 6570543002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if*

   a) any applicant named in part 3 is not an inventor, or

   b) there is an inventor who is not named as an applicant, or

   c) any named applicant is a corporate body.

   See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	12
Claim(s)	4
Abstract	
Drawing(s)	6 + 6

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Lynda M Slack Date 31 July 1997

12. Name and daytime telephone number of person to contact in the United Kingdom Lynda Slack -01625-516173

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

NOVEL COMPOUNDS

This invention relates to a novel human gene (ZGGBP1) associated with affective neurological disorders such as bipolar affective disorder. The invention also relates to homologues of the ZGGBP1 gene in species such as rat and mouse useful in providing animal models of affective disorders. The invention further relates to both the cDNA and the structural gene and to fragments encoding functional domains within the gene. This invention also relates to means for producing the protein encoded by the gene and to means for regulating its production and activity in vivo.

Affective disorders comprise a broad and heterogeneous category of psychiatric illness with a prevalence of up to 20% in the population. The most severe of these disorders is bipolar type I which affects approximately 1% of the population and this rate is fairly consistent across countries. The disease affects young adults, with a mean age of onset of 22 years. Treatment depends upon the phase of the disease and pharmacological agents include lithium carbonate, carbamazepine or valproic acid, tricyclic antidepressants. Monoamine oxidase inhibitors and selective serotonin re-uptake inhibitors are now also being used. The success rate of individual drugs is variable and some patients are treated with a combination of agents, although most have some unwanted side-effects. At present the precise diagnosis of individual affective disorders is difficult and new, gene based, diagnostic methods are desirable.

Family, twin and adoption studies have suggested the importance of genetic predisposition to bipolar affective disorder. On this basis, several groups have undertaken genetic linkage analysis in families with a high incidence of the disorder to find a causal gene. Many of the studies show conflicting data, suggesting that a single gene is unlikely to be the cause. Rather, multiple interacting genetic traits may be involved. A recent study (Stine et al. 1995) identified two regions on chromosome 18 showing linkage to the disease.

The present invention is based on our discovery of a novel gene which maps to 18q21 and which unexpectedly shows appreciable sequence homology to the ned-4 gene on chromosome 15. Ned-4 is the human homologue of the mouse nedd-4 gene which is known

to be differentially expressed during neural development. and to be involved in signal transduction. Human ned-4 has been shown (Schild et al. 1996, Straub et al. 1996) to be a negative regulator of a sodium channel which is deleted in Liddle's syndrome (a hereditary form of hypertension).

Nedd-4 was originally isolated as a partial cDNA clone from a mouse brain library (Kumar et al. 1992) as one of a set of genes which were differentially expressed during development (Neural precursor cells expressed developmentally down-regulated). The derived amino acid sequence contains three copies of the WW domain (Andre & Springael 1994, Bork & Sudol, 1994; Hofmann & Boucher, 1995), a Ca lipid binding (CaLB/C2) domain (Brose et al. 1995) and a Hect (homologous to the E6-AP carbodyl terminus) domain which has homology to a ubiquitin ligase (E3) enzyme (Huibregtse et al. 1995). The human homologue of nedd-4 (Ned-4) was isolated as an randomly cloned EST (KIAA0093) from immature myeloblast mRNA (Nomura et al. 1994) and shown by sequence comparison to have 86% identity at the amino acid level to the mouse sequence. The human sequence, however, has a fourth copy of the WW domain.

The WW domain is a 40 amino acid sequence found in several unrelated proteins. The two highly conserved tryptophans give it its name. The function of the domain is thought to be involved in protein-protein interactions. Despite their functional diversity, the proteins listed all appear to be involved in cell signalling or regulation. It has been shown that the WW domains of Nedd-4 ineract with the proline-rich PY motifs in the epithelial sodium channel in the kidney (Schild et al. 1996). Mutational deletion of the PY motifs in the epitheliam sodium chanel in Liddle's syndrome, an inherited disease causing systemic hypertension characterised by hyperactivity of the sodium channel, has been shown to abrogate binding of Nedd-4 (Straub et al. 1996). It is therefore likely that Nedd-4 has a negative regulatory role when bound to the channel.

The Hect domain is an E3 ubiquitin-protein ligase domain and enzymes with this domain catalyse polyubiquitination, which is involved in several cellular processes including proteolytic degradation, processing protein trafficking.

The CaLB/C2 domain is thought to be involved in calcium-dependent phospholipid binding, although some proteins containing this domain do not bind calcium and other putative functions for the C2 domain such as binding to inositol -1,3,4,5-tetraphosphate have been suggested. Examples of proteins containing this domain are Protein Kinase C (PKC) isoenzymes and synaptogamins.

Therefore in a first aspect of the present invention we provide the ZGGBP1 gene having the cDNA as set out in Figure 1, and fragments thereof. By fragments we mean contiguous regions of the gene including complementary DNA and RNA sequences, starting with short sequences useful as probes or primers of say about 8-50 bases, such as 10-30 bases or 15-35 bases, to longer sequences of up to 50, 100, 200, 500 or 1000 bases. Indeed any convenient fragment of the gene of say up to 2kb, 3kb, 4kb or more than 4kb may be a useful gene fragment for further research, therapeutic or diagnostic purposes. Further convenient fragments include those whose terminii are defined by restriction sites within the gene of one or more kinds, such as any combination of Rsa1, Alu1 and Hinf1.

In a further aspect of the invention we provide homologues of the ZGGBP1 gene in species such as rat and mouse useful in providing animal models of affective disorders. The full sequences of the individual homologues may be determined using conventional techniques such as hybridisation, PCR and sequencing techniques, starting with any convenient part of the sequence set out in Figure 1. The partial sequence of the mouse gene is set out in Figure 4 and this gene and the protein encoded by this gene represent further independent aspects of the invention.

In a further aspect of the invention we provide a recombinant ZGGBP1 protein obtained by expression of all or a part of the cDNA as set out in Figure 1. The recombinant protein may comprise all or a convenient part of the peptide sequence set out in Figure 1. The production of a protein according to the invention may be achieved using standard recombinant DNA techniques involving the expression of the protein by a host cell as described for example by Sambrook et al. 1989. The isolated nucleic acids described herein may for example be introduced into any convenient expression vector for example the T7 Studier system for expression in E.coli (US-A-4952496), Pichia pastoris for expression in

yeast, the Baculovirus system for expression in insect cells and the GS system for expression in mammalian cells by operatively linking the DNA to any necessary expression control elements therein and transforming any suitable procaryotic or eukaryotic host cell with the vector using well known procedures.

Therefore in a further aspect of the invention we provide a recombinant plasmid comprising all or a part of the ZGGBP1 cDNA of the invention.

The invention further extends to cells containing said recombinant plasmids and to a process for producing a ZGGBP1 protein of the invention which comprises culturing said cells such that the desired protein is expressed and recovering the protein from the culture.

By way of example, the nucleotide sequence in Figure 1 is inserted downstream of the SV40 promoter in the pGEX plasmid vector, and either transiently or stably expressed in COS -7 cells. Expression of the protein according to the invention can be detected following disruption of the cells by Western blotting .

It may be desirable to produce the individual functional domains of the protein according to the invention in isolation from the rest of the molecule. This may be achieved using the above standard recombination DNA techniques except that in this instance the DNA sequence used is that encoding one of the partial amino acid sequences of the domains identified in Figure 1 or a combination of these.

By way of further example, the nucleotide sequence in Figure 1 is inserted downstream of the SV40 promoter and the glutathione-S-transferase (GST) coding sequence in the pBC plasmid vector, and either transiently or stably expressed in COS -7 cells allowing expression of the corresponding fusion protein. Expression of the fusion protein can be detected following disruption of the cells by Western blotting with antibodies to GST, and furthermore the fusion protein can be used in an affinity binding procedure to find proteins which are functional partners of the protein of the invention from cell extracts.

A ZGGBP1 protein of the invention may in particular be used to screen for compounds which regulate the activity of the enzymes and the invention extends to such a screen and to the use of compounds obtainable therefrom to regulate the activity of the protein in vivo.

Thus according to a further aspect of the invention we provide a method for identifying a compound capable of modulating the action of a ZGGBP1 protein which method comprises subjecting one or more test compounds to a screen comprising (A) a protein containing the amino acid sequence shown in Figure 1 or a homologue or fragment thereof containing a nucleotide sequence in Figure 1 or a homologue or fragment thereof.

The screen according to the invention may be operated using conventional procedures, for example by bringing the test compound or compounds to be screened and an appropriate substrate into contact with the protein or a cell capable of producing it and determining affinity for the protein in acccordance with conventional procedures.

Any compound identified in this way may be used in the treatment of humans and/or other animals of one or more of the above mentioned diseases. The invention thus extends to a compound selected through its ability to regulate the activity of the protein in vivo as primarily determined in a screening assay utlising the protein containing an amino acid sequence shown in Figure 1 or a homologue or fragment thereof, or a gene coding therefor for use in the treatment of a disease in which the over- or under-activity or unregulated activity of the protein is implicated.

The ZGGBP1 gene of the invention may also be used as the basis for diagnosis, for example to determine expression levels in a human subject, by for example direct DNA sequence comparison or DNA/RNA hybridisation assays. Diagnostic assays may involve the use of nucleic acid amplification technology such as the PCR and in particular the Amplification Refractory Mutation System (ARMS) as claimed in our European Patent No. 0 332 435. Such assays may be used to determine allelic variants of the gene, for example insertions, deletions and/or mutations such as one or more point mutations. Such variants may be heterozygous or homozygous.

The ZGGBP1 gene may also be used in gene therapy, for example where it is desired to modify the production of the protein in vivo, and the invention extends to such uses.

Knowledge of the gene according to the invention also provides the ability to regulate its expression in vivo by for example the use of antisense DNA or RNA. Thus, according to a

further aspect of the invention we provide an antisense DNA or an antisense RNA of a gene for cooperation with the nucleotide sequence shown in Figure 1.

The antisense DNA or RNA for cooperation with the gene in Figure 1 can be produced using conventional means, by standard molecular biology and/or by chemical synthesis as described above. If desired,the antisense DNA or antisense RNA may be chemically modified so as to prevent degradation in vivo or to facilitate passage through a cell membrane and/or a substance capable of inactivating mRNA, for example ribozyme, may be linked thereto and the invention extends to such constructs.

The antisense DNA or antisense RNA may be of use in the treatment of diseases or disorders in humans in which the over- or underregulated production of the gene product has been implicated. Such diseases or disorders may include those described under the general headings of neurologic, eg.stroke, dementia, renal eg.hypertension, nephrosis, cardiovascular disorders.

Convenient DNA sequences may be obtained using conventional molecular biology procedures, for example by probing a human genomic or cDNA library with one or more labelled oligonucleotide probes containing 10 or more contiguous nucleotides designed using the nucleotide sequences described here. Alternatively, pairs of oligonucleotides one of which is homologous to the sense strand and one to the antisense strand, designed using the nucleotide sequences described here to flank a specific region of DNA may be used to amplify that DNA from a cDNA library.

The ZGGBP1 protein of the invention and homologues or fragments thereof may be used to generate substances which selectively bind to it and in so doing regulate the activity of the enzymes. Such substances include, for example, antibodies, and the invention extends in particualr to an antibody which is capable of recognising one or more epitopes containing the protein binding domains shown in Figure 1. In particular the antibody may be neutralising antibody.

As used herein the term antibody is to be understood to mean a whole antibody or a fragment thereof, for example a F(ab)2, Fab, FV,. VH or VK fragment, a single chain

antibody, a multimeric monospecific antibody or fragment thereof, or a bi- or multi-specific antibody or fragment thereof.

The invention will now be illustrated but not limited by reference to the following detailed description, References, Examples and Figures wherein:

Figure 1 shows the sequence of the ZGGBP1 cDNA and predicted amino acid translation. The C2 domain is indicated by carets, the four WW domains are indicated by asterisks and the Hect domain is indicated by underlining.

Figure 2 shows a comparison of amino acid sequences of human ned4 Swissprot entry P46934 and ZGGBP1.

Figure 3 shows a Northern blot analysis of various human tissues probed with ZGGBP1.

Figure 4 shows a comparison of the nucleic acid sequences of human and mouse ZZGBP1. The mouse sequence is a partial cDNA which spans the C-terminal portion of the human protein coding region.

References

Stine O.C. et al. (1994) Am.J.Hum.Genet. *57*, 1384-1394;

Lovett M., Kere J. and Hinton L. (1991) Proc.Natl.Acad.Sci. USA *88*, 9628-9632;

Schuler G.D. et al. (1996) Science *274*, 540-546;

Kumar S., Tomooka Y. and Noda M. (1992) Biochem.Biophys.Res.Commun. *185*, 1155-1161;

Andre B. and Springael J.Y. (1994) Biochem.Biophys.Res.Commun. *205*, 1201-1205;

Bork P. and Sudol M. (1994) Trends Biochem.Sci. *19*, 531-533;

Hofmann K. and Boucher P. (1995) FEBS Letts. *358*, 153-157;

Brose N., Hofmann K.O., Hata Y., Suedhof T.C.(1995) J. Biol. Chem. *270*, 25273-25280;

Huibregtse J.M., Scheffner M., Beaudenon S. and Howley P.M. (1995) Proc.Natl.Acad.Sci. USA, *92*, 2563-2567;

Nomura N. et al. (1994) DNA Res. *2*, 37-43;

Schild L., Lu Y., Gautschi I., Schneeberger E., Lifton R.P. and Rossier B.C. (1996) EMBO J. *15*, 2381-2387;

Straub O., Dho S., Henry P.C., Correa J., Ishikawa T., McGlade J. and Rotin D. (1996) EMBO J. *15* , 2371-2380;

Shizua H., Birren B., Kim U-J, Mancino V., Slepak T., Tachiiri Y. and Simon M. (1992) Proc.Natl.Acad.Sci. USA *89*, 8794-8797;

Kim U-J., Shizuya H., Kang H-L., Choi S-S., Garrett C.L., Smink L.J., Birren B.W., Korenberg J.R., Dunham I. and Simon M. (1996) Proc.Natl.Acad.Sci. USA *93*, 6297-6301.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) "Molecular Cloning: A Laboratory Manual (2nd edition)" Cold Spring Harbor Laboratory Press NY

Example 1

Identification of ZGGBP1

We used two methods for investigating the 18q21 region of interest. In one method we used positional cloning to identify novel transcripts from physical clones representing the region and in a second method we utilised public databases to identify transcripts which had been assigned to a low resolution map of the region by radiation hybrid mapping and assigned them to physical clones representing a high resolution map of the region.

Method 1 - Positional Cloning

The 18q21 region described by Stine et al. (1995) is delimited by the STS markers used by that group to identify linkage. They found the most strongly linked marker to be D18S41, which had a LOD score of 3.51 in cases of paternal inheritance. Linkage declined over flanking markers. We identified a set of four Yeast Artificial Chromosomes (YACs) which comprised a contiguous overlapping set of genomic clones covering the defined region by the presence in those YACs of STS markers used in the Stine study.

DNA from the YACs was prepared and used in a PCR-based hybridisation approach to enrich for transcripts from a human fetal brain cDNA library. This approach, known as

direct selection (Lovett et al. 1991) has been shown to be efficient in identifying transcripts present on large genomic clones.

Method 2 - Refining Radiation Hybrid Mapped Transcripts

The UNIGENE database is a repository for transcripts which have been mapped by taking representative Expressed Sequence Tagged Sites (ESTs) and performing PCR analysis on a panel of radiation hybrids which have been calibrated with respect to a framework of 1000 genetic markers (Schuler et al. 1996). We found 36 EST clusters which had been mapped to a radiation hybrid map interval which corresponded to the 18q21 region of interest and to flanking regions outside.

All the ESTs were tested by PCR on our YAC genomic clones to determine which were present. We found approximately half of the ESTs to be present within the genomic clones and were able to order them based on their position within the YAC contig.

Results

Several clones from our direct selection experiments showed sequence homology to a known EST which we had previously shown to be present in two of the YACs within the contig. The EST was representative of a cluster of sequences. All of these sequences were assembled together using DNAStar Seqman and the consensus sequences obtained were used iteratively to search for other database members within both Unigene, dbEST and EMBL databases. This resulted in the surprising identification of two further clusters of ESTs which had previously not been related to each other on the basis of sequence analysis. The two new EST clusters were annotated as having sequence similarity to ned-4. This was an unexpected finding since we had recently mapped the human ned-4 by Fluorescence In Situ Hybridisation (FISH) to chromosome 15. We were aware that ned-4 was involved in neuronal cell signalling and we concluded that the EST cluster on 18q21 must represent a closely related gene and therefore likely to be involved in affective neurological disorders such as bipolar affective disorder.

The assembly of the EST clusters did not give rise to a single complete contiguous sequence. The reason for this is that many of the EST sequences were derived from IMAGE cDNA clones for which end sequence only was available. In order to fill in the gaps and give a complete contig, four of these clones (IMAGE I.D. 80951, 33059, 79526 and 79984) were sequenced completely to fill the gaps and give an entire complete contiguous sequence. Comparison of the sequence with ned-4 showed that the contig comprised 2kb of 3'Untranslated Region (UTR) and 700bp of the coding region of a gene which had approximately 85% identity at the amino acid level to ned-4 and which we named ZGGBP1.

Isolation of the full length gene for ZGGBP1

The exending of partial transcripts to full length clones can be a complex and difficult process requiring skill and expertise for success. Having considered several possibilities, we opted for a PCR-based approach to isolate and characterise the full length ZGGBP1 gene. Human foetal brain double stranded cDNA was synthesised from mRNA using standard methods (Sambrook et al. 1989) and ligated into lambda Zap vector by use of adapters. However, in order to minimise the loss of transcripts often seen following the cloning step, the resulting ligation mix was not cloned but was instead used as a template for PCR. Oligonucleotide primers specific to ZGGBP1 were used in combination with vector specific primers to amplify DNA across the unknown part of the gene. Since the distance to be covered was unknown, we performed long PCR using the commercially available BCL Expand enzyme and long (30mer) oligonucleotide primers. Since we were using unamplified material, where our target cDNAs were likely to be present only in very small amounts, we utilised a secondary PCR step with nested oligonucleotide primers and again using long PCR to yield sufficient PCR products to be visible by gel analysis and also to minimise the possibility of non-specific PCR amplification. The PCR products derived from these experiments were then purified and sequenced directly. Where necessary, the DNA sequence obtained was used to design further primers to walk along the gene in a 3' - 5' direction. The complete nucleotide sequence derived from this work is 4.8kb and the translated amino acid sequence is shown in Figure 1.

The amino acid sequence derived from the cDNA was compared with that of ned-4 and is shown in Figure 2. The proteins diverge markedly towards the N-terminal portion of the protein, although there is conservation of the common functional motifs.

Northern analysis using a probe derived from the 3'UTR of ZGGBP1 showed a band at approximately 4.8kb but also a more abundant band of 9kb in size in several neurological tissues, with the exception of medulla or spinal cord. These bands are likely to be due to alternative splicing (Figure 3). Other tissues contained the 4.8kb band at higher abundance with respect to the 9kb band and also a 4kb band. ZGGBP1 was expressed in all tissues examined with the exception of liver where we could not detect a transcript at our current detection sensitivity.

Comparison of Amino Acid Sequences of human ned-4 and ZGGBP1

A comparison of the amino acid sequences of human ned-4 and ZGGBP1 is shown in Figure 6. The two proteins have a high level of homology over much of the C-terminal region, including the Hect and WW domains, but diverge over the central portion of the protein. There is a further block of homology near to the N-terminal region, including the C2 domain. The presence of these domains in ZGGBP1 suggests some common functionality with ned-4.

Isolation of Genomic Clone for ZGGBP1

The Research Genetics human Bacterial Artificial Chromosome (BAC) library (Shizua et al. 1992, Kim et al. 1996) was screened by PCR using primers specific to the 3'UTR of ZGGBP1 and BACs were isolated. These will be used to characterise the structural gene including the intron/exon structure and the 5' regulatory region.

Isolation of Mouse homologue for ZGGBP1

The full length sequence of ZGGBP1 shown in Figure 1 was used to search the dbEST database to identify homologous mouse sequences. Three overlapping IMAGE clones were identified (IMAGE I.D.479436, 573510, 482922) comprising a partial transcript. Comparison

of the mouse and human nucleotide sequence is shown in Figure 4. The mouse clones were isolated for use as a probe for in situ hybridisation on sections of mouse brain during development, and as a probe of mouse genomic libraries to isolate genomic clones and to produce transgenic mice by gene targeting using homologous recombination.

AP97-043

NGAP/KEB: 31JUL97

FIGURE 1

Sequence of ZGGBP1 cDNA and predicted amino acid translation

```
1    TCG ACC GCG GCC GCG TCG ACG ACA GGC TCG CGT TAT GGC TCC GCT TTC TGC GGC TCT CCT
1    Ser Thr Ala Ala Ala Ser Thr Thr Gly Ser Arg Tyr Gly Ser Ala Phe Cys Gly Ser Pro

61   ACC CTG GCA TGG TGT GTG TGT GTG CCT GTG TGC TAC GGA GAG TCC CGT ATT CTC AGA GTA
21   Thr Leu Ala Trp Cys Val Cys Val Pro Val Cys Tyr Gly Glu Ser Arg Ile Leu Arg Val

121  AAA GTT GTY TCT GGA ATW GAT CTC GCC AAA AAG GAC ATC TTT GGA GCC AGT GAT CCG TAT
41   Lys Val Val Ser Gly Ile Asp Leu Ala Lys Lys Asp Ile Phe Gly Ala Ser Asp Pro Tyr

181  GTG AAA CTT TCA TTG TAC GTA GCG GAT GAG AAT AGA GAA CTT GCT TTG GTC CAG ACA AAA
61   Val Lys Leu Ser Leu Tyr Val Ala Asp Glu Asn Arg Glu Leu Ala Leu Val Gln Thr Lys
     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

241  ACA ATT AAA AAG ACA CTG AAC CCA AAA TGG AAT GAA GAA TTT TAT TTC AGG GTA AAC CCA
81   Thr Ile Lys Lys Thr Leu Asn Pro Lys Trp Asn Glu Glu Phe Tyr Phe Arg Val Asn Pro
     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

301  TCT AAT CAC AGA CTC CTA TTT GAA GTA TTT GAC GAA AAT AGA CTG ACA CGA GAC GAC TTC
101  Ser Asn His Arg Leu Leu Phe Glu Val Phe Asp Glu Asn Arg Leu Thr Arg Asp Asp Phe
     ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

361  CTG GGC CAG GTG GAC GTG CCC CTT AGT CAC CTT CCG ACA GAA GAT CCA ACC ATG GAG CGA
121  Leu Gly Gln Val Asp Val Pro Leu Ser His Leu Pro Thr Glu Asp Pro Thr Met Glu Arg
     ^^^^^^^^^^^^^^^^

421  CCC TAT ACA TTT AAG GAC TTT CTC CTC AGA CCA AGA AGT CAT AAG TCT CGA GTT AAG GGA
141  Pro Tyr Thr Phe Lys Asp Phe Leu Leu Arg Pro Arg Ser His Lys Ser Arg Val Lys Gly

481  TTT TTG CGA TTG AAA TGG CCT ATA TGC CAA AAA AAT GGA GGT CAA GAT GAA GAA AAC AGT
161  Phe Leu Arg Leu Lys Trp Pro Ile Cys Gln Lys Asn Gly Gly Gln Asp Glu Glu Asn Ser

541  GAC CAG AGG GAT GAC ATG GAG CAT GGA TGG GAA GTT GTT GAC TCA AAT GAC TCG GCT TCT
181  Asp Gln Arg Asp Asp Met Glu His Gly Trp Glu Val Val Asp Ser Asn Asp Ser Ala Ser

601  CAG CAC CAA GAG GAA CTT CCT CCT CCT CCT CTG CCT CCC GGG TGG GAA GAA AAA GTG GAC
201  Gln His Gln Glu Glu Leu Pro Pro Pro Pro Leu Pro Pro Gly Trp Glu Glu Lys Val Asp
                                                  ..................................

661  AAT TTA GGC CGA ACT TAC TAT GTC AAC CAC AAC AAC CGG ACC ACT CAG TGG CAC AGA CCA
221  Asn Leu Gly Arg Thr Tyr Tyr Val Asn His Asn Asn Arg Thr Thr Gln Trp His Arg Pro
     ..................................................................................

721  AGC CTG ATG GAC GTG TCC TCG GAG TCG GAC AAT AAC ATC AGA CAG ATC AAC CAG GAG GCA
241  Ser Leu Met Asp Val Ser Ser Glu Ser Asp Asn Asn Ile Arg Gln Ile Asn Gln Glu Ala
     ........

781  GCA CAC CGG CGC TTC CGC TCC CGC AGG CAC ATC AGC GAA GAC TTG GAG CCC GAG CCC TCG
261  Ala His Arg Arg Phe Arg Ser Arg Arg His Ile Ser Glu Asp Leu Glu Pro Glu Pro Ser

841  GAG GGC GGG GAT GTC CCC GAG CCT TGG GAG ACC ATT TCA GAG GAA GTG AAT ATC GCT GGA
281  Glu Gly Gly Asp Val Pro Glu Pro Trp Glu Thr Ile Ser Glu Glu Val Asn Ile Ala Gly

901  GAC TCT CTC GGT CTG GTT TTG CCC CCA CCA CCA GCC TCC CCA GGA TCT CGG ACC AGC CCT
301  Asp Ser Leu Gly Leu Val Leu Pro Pro Pro Pro Ala Ser Pro Gly Ser Arg Thr Ser Pro

961  CAG GAG CTG TCA GAG GAA CTA AGC AGA AGG CTT CAG ATC ACT CCA GAC TCC AAT GGG GAA
321  Gln Glu Leu Ser Glu Glu Leu Ser Arg Arg Leu Gln Ile Thr Pro Asp Ser Asn Gly Glu

1021 CAG TTC AGC TCT TTG ATT CAA AGA GAA CCC TCC TCA AGG TTG AGG TCA TGC AGT GTC ACC
341  Gln Phe Ser Ser Leu Ile Gln Arg Glu Pro Ser Ser Arg Leu Arg Ser Cys Ser Val Thr

1081 GAC GCA GTT GCA GAA CAG GGC CAT CTA CCA CCG CCA TCA GTG GCC TAT GTA CAT ACC ACG
361  Asp Ala Val Ala Glu Gln Gly His Leu Pro Pro Pro Ser Val Ala Tyr Val His Thr Thr

1141 CCG GGT CTG CCT TCA GGC TGG GAA GAA AGA AAA GAT GCT AAG GGG CGC ACA TAC TAT GTC
381  Pro Gly Leu Pro Ser Gly Trp Glu Glu Arg Lys Asp Ala Lys Gly Arg Thr Tyr Tyr Val
     ..................................................................................

1201 AAT CAT AAC AAT CGA ACC ACA ACT TGG ACT CGA CCT ATC ATG CAG CTT GCA GAA GAT GGT
401  Asn His Asn Asn Arg Thr Thr Thr Trp Thr Arg Pro Ile Met Gln Leu Ala Glu Asp Gly
     ..................................................

1261 GCG TCC GGA TCA GCC ACA AAC AGT AAC AAC CAT CTA ATC GAG CCT CAG ATC CGC CGG CCT
421  Ala Ser Gly Ser Ala Thr Asn Ser Asn Asn His Leu Ile Glu Pro Gln Ile Arg Arg Pro

1321 CGT AGC CTC AGC TCG CCA ACA GTA ACT TTA TYT GCC CCG CTG GAG GGT GCC AAG GAC TCA
441  Arg Ser Leu Ser Ser Pro Thr Val Thr Leu Xxx Ala Pro Leu Glu Gly Ala Lys Asp Ser

1381 CCC GTA CGT CGG GCT GTG AAA GAC ACC CTT TCC AAC CCA CAG TCC CCA CAG CCA TCA CCT
461  Pro Val Arg Arg Ala Val Lys Asp Thr Leu Ser Asn Pro Gln Ser Pro Gln Pro Ser Pro

1441 TAC AAC TCC CCC AAA CCA CAA CAC AAA GTC ACA CAG AGC TTC TTG CCA CCC GGC TGG GAA
```

FIGURE 1 (continued)

```
481  Tyr Asn Ser Pro Lys Pro Gln His Lys Val Thr Gln Ser Phe Leu Pro Pro Gly Trp Glu

1501 ATG TGG ATA GCG CCA AAC GGC CGG CCC TTC TTC ATT GAT CAT AAC ACA AAG ACT ACA ACC
501  Met Trp Ile Ala Pro Asn Gly Arg Pro Phe Phe Ile Asp His Asn Thr Lys Thr Thr Thr

1561 TGG GAA GAT CCA CGT TTG AAA TTT CCA GTA CAT ATG CGG TCA AAG ACA TTT TTA AAC CCC
521  Trp Glu Asp Pro Arg Leu Lys Phe Pro Val His Met Arg Ser Lys Thr Phe Leu Asn Pro

1621 AAT GAC CTT GGC CCC CTT CCT CCT GGC TGG GAA GAA AGA ATT CAC TTG GAT GGC CGA ACG
541  Asn Asp Leu Gly Pro Leu Pro Pro Gly Trp Glu Glu Arg Ile His Leu Asp Gly Arg Thr

1681 TTT TAT ATT GAT CAT AAT AGC AAA ATT ACT CAG TGG GAA GAC CCA AGA CTG CAG AAC CCA
561  Phe Tyr Ile Asp His Asn Ser Lys Ile Thr Gln Trp Glu Asp Pro Arg Leu Gln Asn Pro

1741 GCT ATT ACT GGT CCG GCT GTC CCT TAC TCC AGA GAA TTT AAG CAG AAA TAT GAC TAC TTC
581  Ala Ile Thr Gly Pro Ala Val Pro Tyr Ser Arg Glu Phe Lys Gln Lys Tyr Asp Tyr Phe

1801 AGG AAG AAA TTA AAG AAA CCT GCT GAT ATC CCC AAT AGG TTT GAA ATG AAA CTT CAC AGA
601  Arg Lys Lys Leu Lys Lys Pro Ala Asp Ile Pro Asn Arg Phe Glu Met Lys Leu His Arg

1861 AAT AAC ATA TTT GAA GAG TCC TAT CGG AGA ATT ATG TCC GTG AAA AGA CCA GAT GTC CTA
621  Asn Asn Ile Phe Glu Glu Ser Tyr Arg Arg Ile Met Ser Val Lys Arg Pro Asp Val Leu

1921 AAA GCT AGA CTG TGG ATT GAG TTT GAA TCA GAG AAA GGT CTT GAC TAT GGG GGT GTG GCC
641  Lys Ala Arg Leu Trp Ile Glu Phe Glu Ser Glu Lys Gly Leu Asp Tyr Gly Gly Val Ala

1981 AGA GAA TGG TTC TTC TTA CTG TCC AAA GAG ATG TTC AAC CCC TAC TAC GGC CTC TTT GAG
661  Arg Glu Trp Phe Phe Leu Leu Ser Lys Glu Met Phe Asn Pro Tyr Tyr Gly Leu Phe Glu

2041 TAC TCT GCC ACG GAC AAC TAC ACC CTT CAG ATC AAC CCT AAT TCA GGC CTC TGT AAT GAG
681  Tyr Ser Ala Thr Asp Asn Tyr Thr Leu Gln Ile Asn Pro Asn Ser Gly Leu Cys Asn Glu

2101 GAT CAT TTG TCC TAC TTC ACT TTT ATT GGA AGA GTT GCT GGT CTG GCC GTA TTT CAT GGG
701  Asp His Leu Ser Tyr Phe Thr Phe Ile Gly Arg Val Ala Gly Leu Ala Val Phe His Gly

2161 AAG CTC TTA GAT GGT TTC TTC ATT AGA CCA TTT TAC AAG ATG ATG TTG GGA AAG CAG ATA
721  Lys Leu Leu Asp Gly Phe Phe Ile Arg Pro Phe Tyr Lys Met Met Leu Gly Lys Gln Ile

2221 ACC CTG AAT GAC ATG GAA TCT GTG GAT AGT GAA TAT TAC AAC TCT TTG AAA TGG ATC CTG
741  Thr Leu Asn Asp Met Glu Ser Val Asp Ser Glu Tyr Tyr Asn Ser Leu Lys Trp Ile Leu

2281 GAG AAT GAC CCT ACT GAG CTG GAC CTC ATG TTC TGC ATA GAC GAA GAA AAC TTT GGA CAG
761  Glu Asn Asp Pro Thr Glu Leu Asp Leu Met Phe Cys Ile Asp Glu Glu Asn Phe Gly Gln

2341 ACA TAT CAA GTG GAT TTG AAG CCC AAT GGG TCA GAA ATA ATG GTC ACA AAT GAA AAC AAA
781  Thr Tyr Gln Val Asp Leu Lys Pro Asn Gly Ser Glu Ile Met Val Thr Asn Glu Asn Lys

2401 AGG GAA TAT ATC GAC TTA GTC ATC CAG TGG AGA TTT GTG AAC AGG GTC CAG AAG CAG ATG
801  Arg Glu Tyr Ile Asp Leu Val Ile Gln Trp Arg Phe Val Asn Arg Val Gln Lys Gln Met

2461 AAC GCC TTC TTG GAG GGA TTC ACA GAA CTA CTT CCT ATT GAT TTG ATT AAA ATT TTT GAT
821  Asn Ala Phe Leu Glu Gly Phe Thr Glu Leu Leu Pro Ile Asp Leu Ile Lys Ile Phe Asp

2521 GAA AAT GAG CTG GAG TTG CTC ATG TGC GGC CTC GGT GAT GTG GAT GTG AAT GAC TGG AGA
841  Glu Asn Glu Leu Glu Leu Leu Met Cys Gly Leu Gly Asp Val Asp Val Asn Asp Trp Arg

2581 CAG CAT TCT ATT TAC AAG AAC GGC TAC TGC CCA AAC CAC CCC GTC ATT CAG TGG TTC TGG
861  Gln His Ser Ile Tyr Lys Asn Gly Tyr Cys Pro Asn His Pro Val Ile Gln Trp Phe Trp

2641 AAG GCT GTG CTA CTC ATG GAC GCC GAA AAG CGT ATC CGG TTA CTG CAG TTT GTC ACA GGG
881  Lys Ala Val Leu Leu Met Asp Ala Glu Lys Arg Ile Arg Leu Leu Gln Phe Val Thr Gly

2701 ACA TCG CGA GTA CCT ATG AAT GGA TTT GCC GAA CTT TAT GGT TCC AAT GGT CCT CAG CTG
901  Thr Ser Arg Val Pro Met Asn Gly Phe Ala Glu Leu Tyr Gly Ser Asn Gly Pro Gln Leu

2761 TTT ACA ATA GAG CAA TGG GGC AGT CCT GAG AAA CTG CCC AGA GCT CAC ACA TGC TTT AAT
921  Phe Thr Ile Glu Gln Trp Gly Ser Pro Glu Lys Leu Pro Arg Ala His Thr Cys Phe Asn

2821 CGC CTT GAC TTA CCT CCA TAT GAA ACC TTT GAA GAT TTA CGA GAG AAA CTT CTC ATG GCC
941  Arg Leu Asp Leu Pro Pro Tyr Glu Thr Phe Glu Asp Leu Arg Glu Lys Leu Leu Met Ala

2881 GTG GAA AAT GCT CAA GGA TTT GAA GGG GTG GAT TAA gca ccc tgt gct tcg ggg gtg gtt
961  Val Glu Asn Ala Gln Gly Phe Glu Gly Val Asp •••

2941 gtt ctt cag gca agt tct gct tgc act ttt gca ttt gcc taa cag act ttt gca gag gcg
3001 atg gca gag agc agc tgc agg cat ggt ccc tgg agc cga gcc ttc acc acg cac tcg tcc
```

FIGURE 1 (continued)

```
3061 aag ttc ggg atg cgg gaa cct ggt ccc agc ttg agt tcc tgc ctt tcc cac cac aaa tta
3121 tca act ggt tga tgt gta cac taa tta cat ttc agg agg act taa tgc tat tta tgt tgt
3181 cct ctg cag gca aag ccc tta ata aat att tta cat cct ttc taa tga caa tga atg gaa
3241 tta atc act caa cag gta tag tat tac gac tca tgt tta ctt ttt aaa atg att tag acc
3301 gat ttt cag att tta ttt cgt tat gat taa aga tgt ctc atg tac ttg gaa aag tga gca
3361 ttt ttt ttt ttt ttt gta ttt cac ttt cat acc agg ctt aat gtc aat gac att ttt att
3421 ttt gaa gta ctc tga cac ctc cac cct cta ctt tat tag aat tgg aag gca aat ttt tgt
3481 cca aaa acc tac aga caa gta ctt tga gag aat ttc caa tat aat att aga cat aat gat
3541 aat ttt ttc cat act cag aat gaa aaa ctg gat att acg ttt ttg ttt tgg ggt ttt ttt
3601 gta caa att tag cta ata gct aca ggc tga gag aat tgt aac ata gca tga caa att ttg
3661 tgt tga ctt gaa agg aat cac acc att att cct tag aag taa tta cat gtg ttc taa cac
3721 att tga gac agg gtt gga ctc cca ttt ctc atc cga gaa att act taa ccc ttc ctg ggc
3781 gct gta cag tca tct ttt att cta ttt cct ctt tgc tgt ttg tag tag aga cat ttt gaa
3841 tga aac ttg gca ctg ctt gat tca aaa ctg tgg aaa cca gat ctg ttt agt ctc ctg ttt
3901 gta tgc gtt tgc taa tgg tag cta aat aac cag ttt ttg ttg taa atg cac caa ttc tga
3961 agg cac ttt atg tac tac atg gag gtc ata tct ggt ttt gtt ttt att ttt tta tca tga
4021 aca tta aat gtg atg atg att tct ttt ccc tgc aca cat ctt tcc ggt gca ata tct atc
4081 aat tgt gaa tct ggc tgc tgg tgt ata aaa acc tgg atg taa agc tga gcc tac aga cct
4141 gtc ctc acc aac tgt ttt gtg att tct act caa cta caa aga ttt att taa tgt act ctt
4201 aat cta act gag ttt tgt tac caa tga cct gtt gca tgc ttc aat acc gtg tac tgc ctg
4261 agt tgt gcc tct tgt gtg cta gat taa aag tga gac aga gac ttg act tga tcc tct gag
4321 ctc aag cta ttg agc tgg tat tgg cax agg act gag ggt acc tgc aca gtt tga ttc ttt
4381 tcc cac gtt gta agt ctc cat tgc aga att gtc gtg ttt gag aaa aca cct gag gca gtg
4441 tgg gag ttg aac gac cct gct gtc cyt ttt aac ctg tgt tgt cct aga ccc tkk tcg ggg
4501 sca gtc agg gga cac cta gag att tga tct cat gcg agt cat caa tag gac aaa aaa gtt
4561 gtg gtt tgg gga ggt ctg ttt gtt aca taa aaa gga cct ttc ggt gta aga aat tgc cgt
4621 ttt tac cct gcc ctg gct ggc atg tga gaa gcc atg gaa ggt tgt ggt tgt aaa tga gtt
4681 gtc taa agg ggt gca gag gcc tga ggt ttc taa aag aag gta gat ttc tac aga gct gag
4741 tgt tgg ttc ctt ttt ctt att ggt tga aaa tta cct ggt agt gat cag aaa act tag atg
4801 cta tgt aac t
```

FIGURE 2

Amino Acid Sequence Alignment of Human NED-4 (SwissProt Accession No.P46934) and ZGGBP1 using Clustal method with PAM250 residue weight table

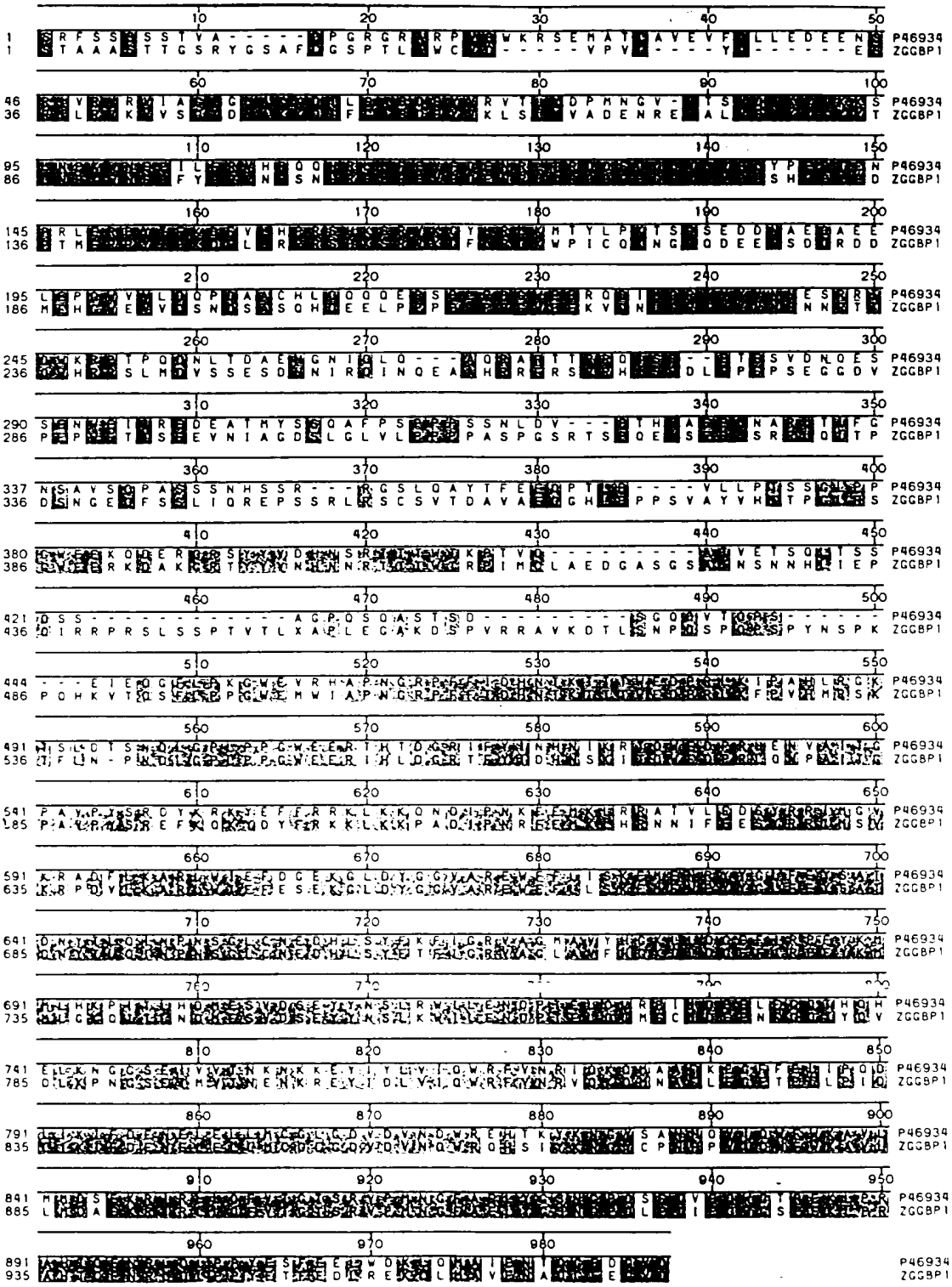

FIGURE 3

Northern Blot Analysis of mRNAs Hybridised with ZGGBP1

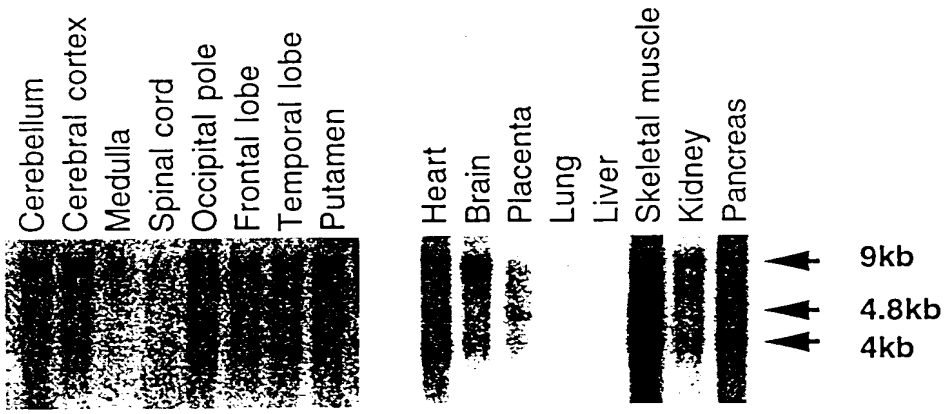

FIGURE 4

Nucleic Acid Sequence Alignment of Human and Mouse ZGGBP1 using Clustal method with Weighted residue weight table